High-frame-rate Full-vocal-tract Imaging based on the Partial Separability Model and Volumetric Navigation

Maojing Fu[1,2], Bo Zhao[1,2], Joseph Holtrop[2,3], Jamie Perry[4], David Kuehn[5], Zhi-Pei Liang[1,2], and Bradley Sutton[2,3]

[1]*Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States,* [2]*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States,* [3]*Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States,* [4]*Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina, United States,* [5]*Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States*

INTRODUCTION

Dynamic MRI holds great potential for visualizing soft tissue movement in the oropharyngeal tract on arbitrary view planes [1]. However, intrinsic trade-offs in spatial resolution, temporal resolution and spatial coverage have limited its use. Precise characterization of the acoustical properties of the vocal tract demands three features: 1) high imaging speed to capture fast articulator movement; 2) full-vocal-tract coverage to explore complex soft-tissue structure in 3D; and 3) sufficient signal-to-noise ratio (SNR) at high temporal resolution. Previously a partial separability (PS) model-based method [2 - 4] with a spiral-navigation-based sampling scheme [5, 6] was developed to enable multiplanar dynamic speech imaging. This work implements a 3D volumetric navigation strategy for the PS model-based approach to capture faster dynamics at a frame rate of 102.2 fps. It also allows for a high spatial resolution at 2.2 mm × 2.2 mm × 6.5 mm and a broad spatial coverage spanning the entire vocal tract with 8 slices.

METHODS

The PS model assumes partial separability of the desired image sequence, $I(\mathbf{r},t) = \sum_{\ell=1}^{L} \xi_\ell(\mathbf{r})\phi_\ell(t)$, where $\xi_\ell(\mathbf{r})$ and $\phi_\ell(t)$ represent the l^{th} spatial and the l^{th} temporal basis functions, respectively [2]. This framework allows reconstruction of $I(\mathbf{r},t)$ from highly undersampled data, $d(\mathbf{k},t)$, by enforcing the partial separability and the spatial-spectral sparsity constraints integratively [3]. The desired image sequence is estimated by:

$$\hat{\mathbf{I}} = \arg\min \|\mathbf{d} - \mathrm{E}\mathbf{I}\|_2^2 + \lambda\|\Phi\mathbf{I}\|_1 ,$$

where E is an imaging operator incorporating both sparse sampling and parallel imaging and Φ is a temporal Fourier transform operator.

Correspondingly, the PS model-based data acquisition scheme sparsely samples the $(\mathbf{k},t)$-space to obtain two data sets: a navigator data set with high temporal resolution and an imaging data set with high spatial resolution. Previous work has enabled multislice dynamic speech imaging using a spiral-navigation-based 2D sampling pattern covering 5 interleaved slices at a frame rate of 20 fps each, but was constrained by the tradeoff of spatial coverage with frame rate [6]. The current work optimizes this PS model-based data acquisition scheme with a 3D volumetric navigation strategy to enable simultaneous navigation across multiple spatial slices. Specifically, this 3D navigation strategy applies a cone trajectory to replace separate 2D navigations and effectively increases the imaging frame rate by reducing the overhead spent on switching between navigation planes. It also allows a larger number of imaging planes to be visualized. In addition, navigator data acquired from this strategy has a potential to yield higher SNR.

To demonstrate the advantage of the 3D navigation strategy, experiments were performed on a Siemens Trio 3T scanner with a 12-channel head receiver coil and a 4-channel neck receiver coil. A FLASH sequence integrating a cone navigation acquisition (TE = 0.85 ms) and a Cartesian imaging acquisition (TE = 2.3 ms) was developed to acquire data over a 260 mm × 260 mm × 52 mm FOV covering the entire vocal tract. The cone trajectory was designed by applying a z-direction gradient to an optimized 2D spiral trajectory [5]. During data acquisition, a volunteer subject produced repetitive /loo/-/lee/-/la/-/za/-/na/-/za/ sounds at a natural speaking rate in accordance with the local IRB. Reconstructions had a matrix size of 128 × 128 × 8, a spatial resolution of 2.2 mm × 2.2 mm × 6.5 mm and a frame rate of 102.2 fps (TR = 9.78 ms) for every slice.

RESULTS

Figure 1 exhibits temporal dynamics of the tongue tip from a mid-sagittal view. Motion patterns of each speech segment are depicted with dotted lines. High imaging speed captures the subtle differences in two seemingly identical events: the tongue takes longer time to produce the first /za/ sound than to produce the second one (indicated by arrows). Figure 2 demonstrates 3D tongue motion from a tilted coronal plane placed beneath the alveolar ridge. Broad spatial coverage allows visualization of both the vocal tract closure at /l/ and the ensuing tongue retraction at /ee/ across arbitrary view angle. Figure 3 reveals the formation of a cavity surrounded by the velum and its lateral wall. High spatial resolution and broad spatial coverage allow the capture of a 3D curved surface when the velum is elevated, compared to a straightened surface when the velum is relaxed. Velum curvatures are also provided in solid lines as references.

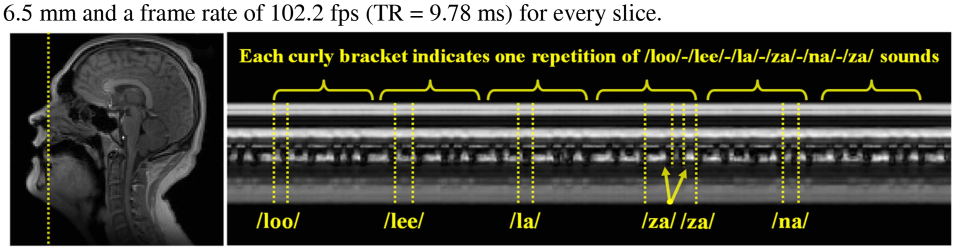

Fig. 1. The temporal dynamics of the tongue tip during the production of /loo/-/lee/-/la/-/za/-/na/-/za/ sounds.

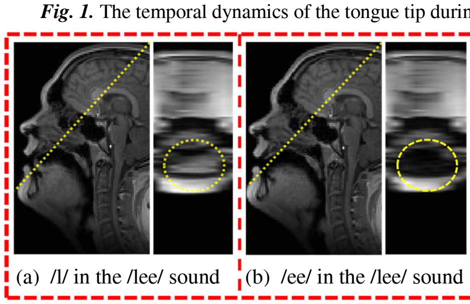

Fig. 2. A tilted coronal plane beneath the alveolar ridge shows (a) the contact of the tongue tip and the alveolar ridge and (b) the ensuing retraction of the tongue tip.

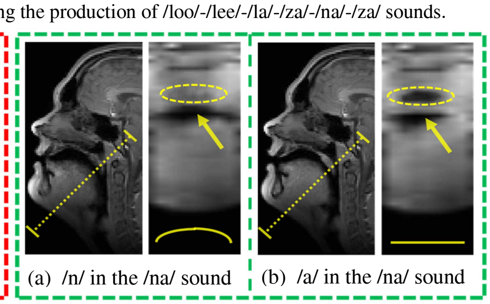

Fig. 3. The velum-pharyngeal-wall cavity (a) closes at the production of the /n/ sound and (b) opens at the /a/ sound. Velum curvatures are given by solid lines.

CONCLUSION

This work enables high-frame-rate dynamic speech imaging based on an optimized 3D navigation scheme. The proposed scheme provides visualization of articulator dynamics over the entire vocal tract with 8 spatial slices at a temporal frame rate of 102.2 fps and a spatial resolution of 2.2 mm × 2.2 mm × 6.5 mm. This method enables dynamic 3D visualization of fast-moving, complex-structure articulators, such as the tongue tip and the velum, and provides a platform for the assessment of the relationship between anatomy and acoustical properties of the oropharyngeal tract.

ACKNOWLEDGEMENT This work was supported by a grant from NIH: 1R03DC009676-01A1 and a CSE fellowship from UIUC.

REFERENCE [1] Niebergall, A. et al. MRM 2012. Preprint. [2] Liang, Z.-P. IEEE-ISBI 2007. pp. 988-991. [3] Zhao, B. et al. IEEE-TMI 2012. pp. 1809-1820. [4] Shosted, R. et al. JASA 2012. p. 2091. [5] Sutton, B. et al. JMRI 2010. pp. 1228-1237. [6] Fu, M. et al. ISMRM 2012. p. 12.